**N-glycolyl groups of nonhuman chondroitin sulfates survive in ancient fossils**

Anne K. Bergfeld, Roger Lawrence, Sandra L. Diaz, Oliver M. T. Pearce, Darius Ghaderi, Pascal Gagneux, Meave G. Leakey, and Ajit Varki

Edited by Chi-Huey Wong, Academia Sinica, Taipei, Taiwan, and approved July 25, 2017 (received for review April 19, 2017)

Biosynthesis of the common mammalian sialic acid N-glycolylneuraminic acid (Neu5Gc) was lost during human evolution due to inactivation of the CMAH gene, possibly expediting divergence of the *Homo* lineage, due to a partial fertility barrier. Neu5Gc catabolism generates N-glycolylhexosamines, which are potential precursors for glycoconjugate biosynthesis. We carried out metabolic labeling experiments and studies of mice with human-like Neu5Gc deficiency to show that Neu5Gc degradation is the metabolic source of UDP-GlcNGc and UDP-GalNGc and the latter allows an unexpectedly selective incorporation of N-glycolyl groups into chondroitin sulfate (CS) over other potential glycoconjugate products. Partially N-glycolylated CS was chemically synthesized as a standard for mass spectrometry to confirm its natural occurrence. Much lower amounts of GalNGc in human CS can apparently be derived from Neu5Gc-containing foods, a finding confirmed by feeding Neu5Gc-rich chow to human-like Neu5Gc-deficient mice. Unlike the case with Neu5Gc, N-glycolyl-CS was also stable enough to be detectable in animal fossils as old as 4 My. This work opens the door for investigating the biological and immunological significance of this glycosaminoglycan modification and for an “ancient glimpses” approach to dating of Neu5Gc loss during the evolution of *Homo*.

All vertebrate cells are covered with a complex array of glycoconjugates, with sialic acids (Sias) typically occupying terminal positions of many glycan chains (1). The two most common Sias in most mammals are N-acetylenuraminic acid (Neu5Ac) and N-glycolylneuraminic acid (Neu5Gc), which differ by a single oxygen atom (1). The only known metabolic pathway for Neu5Gc biosynthesis is hydroxylation of its precursor Neu5Ac, catalyzed by the enzyme CMP-Neu5Ac hydroxylase (CMAH) (2–4). A loss-of-function mutation in the corresponding single-copy *Cmah* gene is fixed in humans, who can no longer biosynthesize Neu5Gc (3, 5, 6). Multiple methods of genomic analysis estimated that the loss of CMAH function occurred ~2–3 Mya in the hominin lineage (7, 8).

There are many known and possible biological consequences of human Neu5Gc loss, discussed in detail elsewhere (9). Following loss of Neu5Gc biosynthesis, the human immune system also recognizes Neu5Gc-bearing glycans as foreign and antigenic molecules (10, 11). In human-like *Cmah*−/− Neu5Gc-deficient mice, the resulting anti-Neu5Gc antibodies could mediate female intrauterine immune selection against sperm from *Cmah*-positive WT males (12, 13). Models of frequency-dependent selection regimes showed that such a reproductive incompatibility by female immunity could drive a loss-of-function allele to fixation after it reached moderate initial frequencies. We therefore postulated that fixation of the CMAH null state in the hominin lineage leading to our species could have expedited divergence of the genus *Homo* ~2–3 Mya (12). However, there is as yet no direct proof for this hypothesis.

Notably, human cellular pathways still allow metabolic incorporation of scavenged Neu5Gc into endogenous cellular glycans (4, 10, 14). This process can even occur in intact humans, by metabolically incorporating trace amounts of Neu5Gc from mammal-derived food products, particularly red meats (beef, lamb, and pork). Incorporation of trace amounts of such exogenous Neu5Gc into human cell-surface structures in the face of an anti-Neu5Gc antibody response makes Neu5Gc the first known “xeno-autoantigen” in humans (15–17). However, incorporation levels are far lower than endogenous levels in WT (*Cmah* intact) mice or in chimpanzees, our closest living evolutionary relatives (18).

As an alternative to activation to CMP-Neu5Gc and incorporation into glycoconjugates, Neu5Gc can also follow a degradative metabolic route, resulting in loss of the N-glycolyl group and formation of glucosamine 6-phosphate (4). One intermediate on this multistep pathway was found to be N-glycolylglucosamine (GlcNGc) (4) (Fig. L4). Additionally, mammalian cells cultured in synthetic GlcNGc were found to biosynthesize UDP-GlcNGc, which was incorporated as O-GlcNGc modifications of proteins (19) and was also detected in O-glycans, likely as N-glycolylgalactosamine (GalNGc) (20, 21). Existence of HexNGCs in O-glycans and N-glycans was also suggested after analyzing cells supplemented with synthetic ManNGc (22). In addition, we found that culturing mammalian cells in synthetic GalNGc gave rise to UDP-GalNGc and UDP-GlcNGc, which serve as sugar nucleotide donors for incorporation into cellular O-glycans, gangliosides, heparan sulfates


Conflict of interest statement: A.V. is cofounder of and has equity interest in SiaMab Therapeutics, Inc., which has licensed University of California, San Diego technologies related to Neu5Gc biology. The terms of this arrangement have been reviewed and approved by the University of California, San Diego in accordance with its conflict of interest policies. However, there are no direct conflicts with the present study.

This article is a PNAS Direct Submission.

1To whom correspondence should be addressed. Email: avarki@ucsd.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1706306114/-/DCSupplemental.
(HS), and chondroitin sulfates (CS) (23). The latter two glycan classes are major components of extracellular matrix and bone (24). While cellular pathways allowed low-level incorporation of artificially provided GalNGc or GlcNGc into most major glycan classes, we noted that the N-acetyl group was most prominently found in CS (23). Here we explore the possibility that metabolic turnover of naturally occurring and/or food-derived Neu5Gc might result in a so-far-unnoticed subset of cellular glycans that naturally contain HexNGc residues. We also apply these findings to challenging questions regarding indirect fossil evidence of Neu5Gc expression over a time span of millions of years.

Results

Possible Metabolic Pathway for Natural Occurrence of N-Glycolyhexosamines in Animal Glycans. Mammalian cells cultured with synthetic N-glycolyhexosamines (HexNGc) incorporate GalNGc and GlcNGc into cellular glycans (19, 22, 23). Meanwhile, ManNGc and GlcNGc were found as natural products generated during metabolic degradation of Neu5Gc (4). Taken together, these findings raised the interesting question of whether cellular turnover of Neu5Gc naturally produces HexNGc that are subsequently incorporated into animal glycans. The proposed metabolic pathway involves the conversion of Neu5Gc via ManNGc and GlcNGc toward GlcNGc-6P (Fig. 1A). As an alternative to loss of the N-glycolyhexosamine (generating glycolate and glucosamine-6-phosphate) (4), GlcNGc-1P might also be formed. Further metabolism toward UDP-GlcNGc and subsequent epimerization to UDP-GalNGc could then provide the precursors for glycan assembly (Fig. 1A). To test this hypothesis, human THP-I cells were cultured with [3H-glycolyl]Neu5Gc followed by isolation of radiolabeled cellular glycoconjugates comprising GlcNGc and GalNGc in nature. (i) Conversion of Neu5Gc into ManNGc is catalyzed by the N-acetylneuraminic lyase (EC 4.1.3.3) (4, 50, 51). (ii) Epimerization of ManNGc to GlcNGc can be achieved by GlcNAc-2′-epimerase (EC 5.1.3.8) (4). (iii) Phosphorylation of GlcNGc in the 6′ position to result in GlcNGc-6-P′s being catalyzed by GlcNAc 6-P phosphomutase (EC 5.4.2.3) (52). (iv) GlcNGc-1-P′ would thereafter react with UTP to form UDP-GlcNGc, a reaction potentially catalyzed by UDP-N-acetylglucosamine diphosphorylase (EC 2.7.7.23) (53). (v) Epimerization of UDP-GlcNGc to UDP-GalNGc is catalyzed by UDP-GlcNAc 4-epimerase (EC 5.1.3.7) (54). UDP-GlcNGc and UDP-GalNGc then serve as precursors for glycan assembly.

Fig. 1. Possible underlying mechanism for the natural occurrence of HexNGc in animal glycoconjugates. (A) The single known source for N-glycolyhexosamines in animals is the conversion of the N-acetyl group of CMP-Neu5Ac to an N-glycolyl group in CMP-Neu5Gc, which is catalyzed by Cmah (EC 1.14.18.2) (49). Therewith, N-glycolyhexosamines are to be Neu5Gc derivatives. Based on the well-studied N-glycolyhexosamine pathways in animals we suggest a metabolic route to result in glycoconjugates comprising GlcNGc and GalNGc in nature. (i) Conversion of Neu5Gc into ManNGc is catalyzed by the N-acetylneuraminic lyase (EC 4.1.3.3) (4, 50, 51). (ii) Epimerization of ManNGc to GlcNGc can be achieved by GlcNAc-2′-epimerase (EC 5.1.3.8) (4). (iii) Phosphorylation of GlcNGc in the 6′ position to result in GlcNGc-6-P′s being catalyzed by GlcNAc 6-P phosphomutase (EC 5.4.2.3) (52). (iv) GlcNGc-1-P′ would thereafter react with UTP to form UDP-GlcNGc, a reaction potentially catalyzed by UDP-N-acetylglucosamine diphosphorylase (EC 2.7.7.23) (53). (v) Epimerization of UDP-GlcNGc to UDP-GalNGc is catalyzed by UDP-GlcNAc 4-epimerase (EC 5.1.3.7) (54). UDP-GlcNGc and UDP-GalNGc then serve as precursors for glycan assembly. (B) Human THP-I cells and (C) CHO lec32.LEC29 cells were cultured with [3H-glycolyl]Neu5Gc. Desalted GAGs were divided into three samples, from which one sample remained untreated (filled gray), one sample was treated with chondroitinase ABC (black line), and the last sample was treated with heparinases (gray line). The disaccharides were separated from the intact GAG chains by gel filtration chromatography.

To confirm the hypothesis, human THP-I cells were cultured with [3H-glycolyl]Neu5Gc followed by isolation of radiolabeled cellular glycoconjugates (GAGs). GAGs were selected because earlier studies indicated that incorporation of synthetic HexNGc in mammalian cells was highest into this class of glycans (23). The GAG fraction contained 2.5% of total cellular radioactivity and was divided into three aliquots, which were treated either with chondroitinase ABC or a mixture of heparin lyases-I/II/III or incubated as a nontreated control. Reaction products were analyzed by size-exclusion chromatography to separate large GAG polymers from disaccharide digestion products. Radioactivity in nontreated GAG chains (shaded gray) eluted early due to their polymeric nature (Fig. 1B). No differences were observed for the heparin lyase-treated sample (gray line), indicating that the glycolyl group derived from [3H-glycolyl]Neu5Gc was not incorporated into HS. In contrast, analysis of the chondroitinase-treated sample (black line) shows a peak eluting later, as expected for CS disaccharides (Fig. 1B), suggesting that the N-[3H-glycolyl]Neu5Gc was incorporated as GalNGc into cellular CS. The same experiment was performed with CHO lec32.LEC29 cells, which lack the ability to activate Sias (25) and might therefore show enhanced metabolic flux of supplemented [3H-glycolyl]Neu5Gc toward HexNGc. Indeed, 8% of cellular radioactivity was now detected in the GAG fraction. Analysis of isolated GAGs was performed as described for...
the human cells (Fig. 1C), and labeled, nontreated polymeric GAGs eluted early as expected (shaded gray). Again, no peak shift was observed after heparin lyase treatment (gray line) and all radioactivity eluted in the disaccharide peak after treatment with chondroitinase. Thus, cells selectively incorporated [3H-glycolyl]GalNGc into CS, a likely location for naturally occurring GalNGc in animal glycans. Notably, despite the fact that UDP-[3H-glycolyl]GalNGc would be the most distal product of [3H-glycolyl]Neu5Gc metabolism (six steps away, see Fig. 1A) it was still preferentially used to synthesize [3H-glycolyl]CS, versus the more proximal product UDP-[3H-glycolyl]GlcNAc, which could have been used for biosynthesis of other glycan classes, such as N-glycans or HS.

**Natural Occurrence of UDP-HexNGc in Livers of WT but Not Neu5Gc-Deficient Mice.** To confirm that Neu5Gc metabolism is indeed the source of naturally occurring HexNGc in animal glycans we compared WT to Cmah−/− mice, which carry the human-like mutation in Cmah and are therefore devoid of any detectable Neu5Gc, making them an excellent negative control (4, 26, 27). Incorporation of HexNGc into animal glycans depends on the generation of its activated UDP-HexNGc precursor. Therefore, nucleoside sugars isolated from Cmah−/− and WT mouse livers were first subjected to liquid chromatography (LC)-MS analyses. In WT mouse liver samples, a peak eluting at 17.7 min (Fig. 2A) with an underlying mass of 606 Da (Fig. 2B) represents UDP-HexNAc sugars. As expected, the same peak was found in corresponding Cmah−/− samples (Fig. 2C and D) at comparable intensity. Samples from WT mice also showed a peak eluting at 17.28 min (Fig. 2E) with an underlying mass of 622 Da (Fig. 2F), which exactly fits the mass of UDP-HexNGc. Complete absence of this peak in samples from Neu5Gc-deficient Cmah−/− mice (Fig. 2G and H) demonstrates that UDP-HexNGc is a natural metabolite derived only from Neu5Gc turnover.

**Chemical Synthesis of N-Glycolyl-CS.** The above data suggest that CS disaccharides from Neu5Gc-containing animal tissues might contain naturally occurring GalNGc. To definitively identify the predicted CS disaccharides containing GalNGc, a synthetic standard was prepared by partial de-N-acetylation and subsequent re-N-glycolylation of commercially available CS from shark cartilage. Shark samples do not have endogenous Neu5Gc (16) and, as expected, commercial shark cartilage CS disaccharides showed no signs of N-glycolyl groups, with product ion spectra of CS disaccharides D0a4 and D0a6 being consistent with published data (for nomenclature, see the references in refs. 23 and 28). Shark CS was partially de-N-acetylated by treatment with hydrazine followed by subsequent re-N-glycolylation with acetoxycetyl chloride (Fig. S1A). The final product was treated with chondroitinase ABC and resulting disaccharides analyzed by glycine reductive isotope labeling (GRIL)-LC/MS as described (23, 28). The extracted ion current is shown for CS disaccharides D0a4 and D0a6 (Fig. S1B) as well as corresponding D0q4 and D0q6 (Fig. S1C) of the partially N-glycolylated product. The nomenclature of N-glycolylated GAGs has been described previously.

![Fig. 2.](Image)
The product ion spectra of the N-glycolylated species D0q4 and D0q6 in Fig. S1 D and E are consistent with published data (23). The ratio of GalNGc:GalNAc in synthesized, partially N-glycolylated CS was estimated to be ~1:10 by MS of disaccharides, assuming similar signal intensities for both substituents. Complete absence of GalNGC in the starting material makes it a suitable negative control, and the ability to generate N-glycolylated CS provides not only the required standard but also the future opportunity to investigate the possible impact and function of N-glycolyl-CS in mammals.

**Natural Occurrence of GalNGC in Animal CS.** Liver GAGs of WT and Cmah−/− mice were isolated and depolymerized by chondroitinase ABC. Released disaccharides were analyzed by GRIL-LC/MS and differentially stable isotope-labeled N-acetylated CS-disaccharides were added as internal standards to all samples (23, 28). Besides the expected N-acetylated CS-disaccharides, samples from WT mouse liver also contained a peak at 22.9 min (35 min run), which fits the N-glycolylated CS-disaccharide D0q4 [m/z = 435 (551 with the aniline tag)] (Fig. 3A). MS-MS analysis of the respective ion by collision-induced dissociation (CID) fits the expected fragmentation pattern for GalNGC-containing D0q4 disaccharide (Fig. 3B) (23). No D0q4-containing CS was detectable in Cmah−/− mouse livers (Fig. 3C). Similarly, N-glycolyl-CS was detectable in WT mouse kidney samples but absent from Cmah−/− mice (Table S1).

After confirming naturally occurring GalNGC using the well-controlled mouse model, samples from additional animal species were analyzed. Given that Neu5Gc was earlier found to be rich in “red meats” (10, 16, 29), we hypothesized that chances to detect its breakdown products are highest in such samples. Therefore, CS from lamb (ovine muscle), beef (bovine muscle), and pork (porcine muscle) were investigated next. The characteristic D0q4 CS-disaccharide peak [m/z = 551] of the lamb sample eluted at 22.8 min (35 min run; Fig. 3D) and its nature was identified by CID analysis, yielding the expected fragmentation pattern of D0q4 as mentioned above (Fig. 3E). Similarly, N-glycolyl-CS was detectable in porcine and bovine muscle (Table S1). In addition, CS purified from FCS and commercially available purified CS from porcine intestinal mucosa and bovine trachea were analyzed. As predicted, N-glycolyl-CS was detectable in all such animal-derived samples (Table S1). Based on the observed ion intensities, it is possible that N-glycolyl-CS exists at levels approaching 0.5%. Such estimates for rare residues in both CS and HS assume similar ionization efficiencies and matrix effects on ion suppression. There is no evidence that a significant ionization deficit exists in going from N-acetyl to N-glycolyl. However, more accurate quantitative work will require the synthesis of higher-purity standards.

**Neu5Gc Ingestion Generates Much Smaller Amounts of GalNGC, Detectable in Human CS.** Given that humans cannot biosynthesize Neu5Gc (5, 6), no GalNGC should be detectable in human CS samples, whereas related chimpanzee samples should contain N-glycolyl-CS. As expected, GalNGC-containing CS-disaccharides are present in chimpanzee serum as analyzed by selective reaction monitoring (SRM) for the predominant D0q4 structure eluting at 22.8 min (35 min run; Fig. 4A). The identity of the respective ion was confirmed by CID analysis (Fig. 4B). Interestingly, purified CS from commercial human serum actually showed trace amounts of N-glycolyl-CS, close to the detection limit of the instrument. The predominant D0q4 structure was eluting at 22.8 min (35 min run; Fig. 4C) and the identity of the ion was verified by MS/MS fragmentation analyses (Fig. 4D). To confirm this finding, multiple additional sera from healthy human volunteers were analyzed individually and trace amounts of GalNGC-containing CS were detectable in the majority of samples (Table S1). Given the complete absence of N-glycolyl groups in samples from Cmah−/− mice fed vegan, soy-based diets, the only conceivable source for GalNGC in humans would be uptake and metabolism of exogenous, diet-derived Neu5Gc.

To substantiate this hypothesis, Neu5Gc-deficient Cmah−/− mice were switched from a vegan soy chow diet onto a soy chow containing Neu5Gc at levels comparable to a human Western diet as described previously (16, 30). As expected, N-glycolyl-CS is readily detectable in WT mouse serum (Fig. 4 E and F) and absent in the sera of vegan Cmah−/− knockout mice (Fig. 4 G and H). As predicted, trace amounts of N-glycolyl-CS were consistently detectable only in Cmah−/− mice on the Neu5Gc-containing diet (Fig. 4 I and J). It has been concluded earlier that humans take up and incorporate the nonhuman, immunogenic Sia Neu5Gc.
Fig. 4. LC/MS analysis of serum from primates and mice. (A and B) The extracted ion current (XIC) for D0q4 in chimpanzee serum identified by retention time and tandem MS (CID spectrum) compared with D0q4 standard. A 35-min gradient was used for this sample. (C and D) The XIC and corresponding CID spectrum for small trace amounts of D0q4 found in human serum. A 35-min gradient was used for this sample. (E and F) The XIC and corresponding CID spectrum for D0q4 found in a WT mouse. A 100-min gradient with higher resolution was used for this sample. (G and H) The XIC and corresponding CID spectrum for the time range where D0q4 normally elutes (arrow) in the 100-min gradient used for this sample. (I and J) The XIC and CID spectrum for D0q4 detected in Cmah−/− mice fed a Neu5Gc-containing diet. A 35-min gradient was used for this sample. The asterisks and double asterisks in the mouse XIC traces are isobaric species that are detected along with D0q4 but do not coelute with D0q4 standard. The asterisks in the CID spectra denote a daughter ion that does not correspond to a product ion from D0q4 standard.
from animal-derived food products (10, 16, 30). The present study suggests that humans also metabolize Neu5Gc toward GalNGc and incorporate this additional nonhuman monosaccharide into their glycoconjugates. However, it is present in trace amounts, much lower than that seen in species with intact CMAH and the capacity to synthesize endogenous Neu5Gc.

Detection of N-Glycolyl-CS in Fossils. Genetic analyses estimate the human loss of Neu5Gc biosynthesis to have occurred ∼2–3 Mya (7, 8). Taken together with mouse fertility studies and models of selection, we could posit that CMAH loss of function may have expedited divergence of the genus Homo, shortly afterward (12). To prove this hypothesis, fossil specimens of early bipedal hominins in the time period 4–2 Mya would ideally need to be analyzed for their Sia composition. Unfortunately, Sias are too unstable to remain detectable in fossil specimens from (sub)tropical regions where most of the more ancient bipedal hominins have been discovered (7). As CS molecules are significantly more stable and also abundant in bones, detection of the Neu5Gc-derivative GalNGc in CS of hominin fossils may allow dating of the human loss of Neu5Gc. For method establishment, CS was isolated from contemporaneous humans and chimpanzees starting with ∼100 mg of powdered bone material. CS was purified from both samples and treated with chondroitinase ABC before analysis by GRIL-LC/MS as described above (Table S2). N-glycolyl-CS was easily detected in chimpanzee bone and much smaller amounts were also noted in human bone, which is in line with the detection of N-glycolyl-CS traces in human sera described above. Thereafter, mammalian fossils (∼12,000–80,000 y old from mammoth, elk, and cave bear) were investigated and CS as well as N-glycolyl-CS was detectable (Table S2). Based on these encouraging findings, we studied 100 mg of bone powder obtained from a ∼1.65-My-old “Java Man” Homo erectus fossil (7, 31). Although CS could be successfully isolated from this significantly older fossil specimen, the total amount of CS in the sample was too low to allow detection of the minor N-glycolylated disaccharide species even if it was present (Table S2).

As the ∼1.6-My-old Java man fossil is very precious and might even be too young to help date the human loss of Neu5Gc, we next focused on 4-My-old Kenyan bovid fossil samples, which were excavated from a bone bed in Area 26l-1 at Allia Bay on the eastern shores of Lake Turkana. Using significantly larger amounts of starting material from these nonprecious samples accompanying famous hominin fossils (5 g each), the presence of N-glycolylated disaccharide residues in all animal fossil samples was evaluated using multiple reaction monitoring for Y1, O2X1, and M-HSO3 daughter ions consistent with N-glycolylated disaccharide residues (m/z = 393, 435 and 471, respectively) for fossil F44. (B) The product ion mass spectrum of D0q4 (m/z value = 551). (C) CS compositional analysis comparing D0q4 and D0a6 abundance of bone and fossil samples expressed as molar percent (mol %).

Fig. 5. Detection of N-glycolyl groups in CS isolated from partially mineralized fossil samples. (A) The accumulative extracted ion current chromatogram for Y1, O2X1, and M-HSO3 daughter ions consistent with N-glycolylated disaccharide residues (m/z = 393, 435 and 471, respectively) for fossil F44. (B) The product ion mass spectrum of D0q4 (m/z value = 551). (C) CS compositional analysis comparing D0a4 and D0a6 abundance of bone and fossil samples expressed as molar percent (mol %).

The altered disaccharide composition in fossil bones likely originates from geochemical processes (biochemical diagenesis) (32) and may be influenced by the individual conditions at excavation sites. Regardless, the unusually different disaccharide composition helps to exclude the possibility that CS or HS detected in fossils results from human contamination during handling and sample preparation. Overall, we successfully demonstrated the presence of CS as well as N-glycolyl-CS in animal fossils as old as 4 My. This method would theoretically allow dating the human loss of Neu5Gc when studying hominin fossils from this era. However, given the extreme rarity of older hominin fossil specimens and the currently needed starting material of 5 g, this can only be followed up when MS instruments reach a higher level of sensitivity.

Discussion

We have demonstrated the natural occurrence of the monosaccharide GalNGc and its activated form UDP-GalNGc, which can act as a donor for animal CS biosynthesis, and proven an
unexpected cellular pathway for metabolism of the \(N\)-glycolyl group from the \(Sia\) Neu5Gc via six metabolic steps, leading to UDP-GalNGc. Conclusive proof of this pathway comes from the absence of UDP-HexNGc and \(N\)-glycolyl-CS in a mouse strain with human-like lack of endogenous Neu5Gc due to inactivation of the \(Cmah\) gene. Further supporting this finding, the presence of GalNGc in CS was shown in tissues and sera of other animals known to be rich in Neu5Gc, including commercially available CS from bovine trachea and porcine intestinal mucosa. In contrast, cartilage CS from shark (another group of species so far found devoid of Neu5Gc) was also devoid of GalNGc in CS. As expected, GalNGc was readily detectable in the chimpanzee specimens.

Considering the multiple metabolic steps involved in generating \(N\)-glycolyl-CS from Neu5Gc, it is surprising that natural occurrence of \(N\)-glycolylo groups is largely limited to CS in vivo. In this regard, it is notable that six metabolic steps tolerate the change from the usual hydrophobic \(N\)-acetyl group to a hydrophilic \(N\)-glycolyl group and thus allow conversion of Neu5Gc to UDP-GalNGc. This apparent “channeling” of the \(N\)-glycolyl group toward CS either means that this low-abundance monosaccharide disturbs cellular pathways at least when incorporated in CS or that \(N\)-glycolylated CS has a specific biological function. Other rare modifications of GAGs such as \(O\)-sulfation of HS are critical for interaction with HS-binding proteins (33, 34). The highly selective enrichment of \(N\)-glycolyl groups only in CS suggests the possibility that \(N\)-glycolyl-CS may have some as-yet-unknown function in mammals that was then also lost in humans. One possible explanation could be the specificity of the metabolic enzymes involved toward HexNAc versus HexNGc. Another possibility would be a shorter lifecycle or possible shedding from cells of other HexNGc-containing glycoconjugates. Regardless, availability of synthetic \(N\)-glycol-lyl-CS will allow future exploration of the functional question, for example by searching for endogenous proteins that selectively bind this GAG.

Unexpectedly, much smaller amounts of GalNGc were also detectable in CS from 8 out of 10 human serum samples as well as in human bone (Table S1). Taken together with the absence in \(Cmah\)−/− mice, the only explanation is that these trace amounts of GalNGc in human CS result from uptake and metabolism of Neu5Gc in animal-derived foods. Substantiating this hypothesis, \(Cmah\)−/− mice fed a Neu5Gc-containing diet showed trace amounts of GalNGc in the CS as well. Direct comparison of human meat and \(Cmah\)−/− mice, the only explanation is that these trace amounts of GalNGc in the CS suggests the possibility that Neu5Gc-null individuals have some as-yet-unknown function in mammals that was then also lost in humans. One possible explanation could be the specificity of the metabolic enzymes involved toward HexNAc versus HexNGc. Another possibility would be a shorter lifecycle or possible shedding from cells of other HexNGc-containing glycoconjugates. Regardless, availability of synthetic \(N\)-glycolyl-CS will allow future exploration of the functional question, for example by searching for endogenous proteins that selectively bind this GAG.

In the present study the chemical synthesis of partially \(N\)-glycolylated CS was developed, established, and fully described. The synthesis is based on commercially available shark cartilage, which is devoid of \(N\)-glycolylated CS and can therefore serve as a negative control. Following partial de-\(N\)-acetylation, partially \(N\)-glycolylated CS could be achieved using acetoxyacetil chloride. The identity of the product was confirmed by MS. This material provides the starting point for comprehensive in vitro studies such as screening for potential binding partners to \(N\)-glycolyl-CS or screening for anti-HexNGc xeno-autoantibodies in the future.

While contemporary humans only have trace amounts of GalNGc in CS (derived from food sources), it is much more easily detectable in chimpanzee samples, as it is derived from the much larger amounts of endogenous Neu5Gc. Thus, analysis of CS from hominin fossils may allow dating the human loss of Neu5Gc biosynthesis. Previous attempts to analyze Sias directly in European Neanderthal specimens were successful (7), but attempts to do so in much older hominin fossils and faunal samples from Africa failed, likely because these sugars are too unstable in specimens from tropical and semitropical regions, where most of the early bipedal hominins have been discovered (35). However, CS is significantly more stable than Sias and provides an alternative option to detect \(N\)-glycolyl groups in ancient fossils. Indeed, GalNGc-containing CS was successfully detected in various faunal fossil samples including a ~4-My-old long bone collected right next to \textit{Australopithecus anamensis} in Allia bay, Kenya (Table S2). As the human loss of Neu5Gc is estimated to have occurred ~2–3 Mya (7) (Fig. 6), detection of GalNGc in hominin fossils from the relevant era should be possible. However, while \(N\)-glycolyl-CS is easily detectable in extracts from 100 mg of contemporary bones, a similar-sized sample obtained from an \(H.\) erectus in Java had very little CS and was below the detection limit for \(N\)-glycolyl-CS. At the current level of sensitivity, we estimate that 5 g of starting material are required for such an analysis, making the study currently unfeasible given the rarity and precious nature of ancient hominin fossils.

However, improved sensitivity of MS and further optimization of CS isolation should render this study possible in the future. At that point it should be possible to access small samples of the relevant hominin fossils, spanning the time period from 1–4 Mya (35, 36), beginning possibly with definitive Homo samples for which larger amounts of material are available [e.g., \textit{Homo georgicus} (37) and \textit{Homo antecessor} (38, 39)]. If indeed such samples are negative for GC-CS, it would then become important to request access to \textit{Australopithecus} and \textit{Paranthropus} samples (35) predicted to be positive, as well as undated samples of lineages with controversial connections to the \textit{Homo} clade such as \textit{Homo naledi} (40, 41) and \textit{Australopithecus sediba} (42, 43). Assuming clear-cut results for samples, the division into \(N\)-glycolyl-CS positive and negative should help clarify the phylogeny of hominin fossils (samples with only trace amounts of \(N\)-glycolyl-CS will be assumed to represent Neu5Gc-null individuals who acquired the traces from dietary Neu5Gc, and thus classified

\begin{figure}
\centering
\includegraphics[width=0.7\textwidth]{figure6.png}
\caption{Potential scenario for the role of Neu5Gc loss and female anti-Neu5Gc immunity in the origin of the genus \textit{Homo} via interplay of natural and sexual selection acting on cell-surface Sias. There are many known pathogens that recognize and exploit Neu5Gc (blue diamond) as a receptor on host target cells (9). Natural selection by such pathogens may have selected for rare \textit{CMAH} null alleles that abolish Neu5Gc expression in homozygote individuals (12). Such individuals have only Neu5SAc and its derivatives on their cells (red diamonds) allowing an escape from pathogens, but at higher frequencies would be targeted by adapting pathogens, resulting in maintenance of a balanced polymorphism. \textit{CMAH}+/− females with anti-Neu5Gc antibodies also present in their reproductive tract would favor sperm from \textit{CMAH}+/− males due to anti-Neu5Gc antibody-mediated cryptic selection against \textit{CMAH}+/− males expressing Neu5Gc on their sperm. Once the frequency of the \textit{CMAH} null allele reaches a critical level, this process can drive fixation of the loss-of-function allele in a population by directional selection (12).}
\end{figure}
additional 50
final concentration of 100 ther. Cells were lysed by adding 2.5 mL of 0.1 M NaOH following 15-min media. The culturing media was supplemented with 7,400 kBq of [3H]Neu5Gc in a total volume of 20 mL. Cells were washed twice with PBS to [3H]Neu5Gc 3 d before reaching confluence. Thereafter, the media was re-

Materials and Methods
Materials. [Glycolyl-3H]Neu5Gc was prepared as described previously (4). Chimpanzee blood samples were collected as extra tubes only during rou-
tine noninvasive health screens of chimpanzee subjects at the Yerkes National Primate Center, Emory University, Atlanta, GA (supported by NIH Base Grant ORIPPO PS10D011132, routine collection covered under local IRB approval by Emory University). All collections were made before the Sep-
tember 15, 2015, designation of captive chimpanzees as endangered species. Human AB serum was purchased from Valley Biomedical. Individual human serum samples were collected from healthy human volunteers, who pro-
vided informed consent. All human samples were completely anonymized,
and the only information available to the investigators was age and sex. All individuals who handled the primary human and primate samples received the required training regarding precautions for blood-borne pathogens. All health and safety issues related to handling of human and nonhuman pri-
mate samples are covered by an Institutional biosafety approval from the University of California, San Diego Environmental Health and Safety Com-
mittee. Commerially available isolated CS from bovine trachea, porcine intestinal mucosa, and shark cartilage were obtained from Sigma. H. erectus samples from Indonesia were kindly provided by Etty Indriati and the late Teuku Jacob, Gadjah Mada University, Yogyakarta, Indonesia, as described earlier (7). Sources of other fossils are indicated in the main text and/or in Table S2.

Tissue Culture. Human acute monocytic leukemia cell line THP-1 (47) was cultivated in DMEM (high glucose; Invitrogen) and CHO LEC29.lec32 cells (25) were cultivated in alpha MEM. Media were supplemented with 5% human serum (heat-inactivated and sterile-filtered; Valley Biomedical Inc.) and 2 mM glutamine. Cells were cultivated in a humidified 5% CO2 atmosphere

Integration of Nucleoside Diphospho Sugars. An adult mouse liver was homog-

Table S2.

Isolation and Analysis of Radiolabeled GAGs from Cells. CHO cells were cul-
tured under the conditions described above in a P-150 dish with 20 mL of media. The culturing media was supplemented with 7,400 kBq of [glycolyl-
3H]Neu5Gc 3 d before reaching confluence. Thereafter, the media was re-

Blood Samples. Chimpanzee blood samples were collected as extra tubes only during routine noninvasive health screens of chimpanzee subjects at the Yerkes National Primate Center, Emory University, Atlanta, GA (supported by NIH Base Grant ORIPPO PS10D011132, routine collection covered under local IRB approval by Emory University). All collections were made before the September 15, 2015, designation of captive chimpanzees as endangered species. Chimpanzee blood samples were shipped overnight on ice to the University of California, San Diego. Human blood was collected at about the other overnight incubation. Thereafter, samples were incubated at 99 °C in a waterbath for 10 min to inactivate the proteinase. Samples were vortexed, spun at 3,000 g for 1 h in 50 mL conical tubes, and supernatant was trans-
ferred to a new tube and spun again at 15,000 g at 3,000 °C for 10 min. Thereafter, samples were incubated at 99 °C in a waterbath for another 10 min to inactivate the proteinase. Samples were vortexed, spun at 3,000 g for 1 h in 50 mL conical tubes, and supernatant was trans-
ferred to a new tube and spun again at 15,000 g at 3,000 °C for 10 min. Thereafter, samples were incubated at 99 °C in a waterbath for another 10 min to inactivate the proteinase. Samples were vortexed, spun at 3,000 g for 1 h in 50 mL conical tubes, and supernatant was trans-
ferred to a new tube and spun again at 15,000 g at 3,000 °C for 10 min. Thereafter, samples were incubated at 99 °C in a waterbath for another 10 min to inactivate the proteinase. Samples were vortexed, spun at 3,000 g for 1 h in 50 mL conical tubes, and supernatant was trans-
ferred to a new tube and spun again at 15,000 g at 3,000 °C for 10 min. Thereafter, samples were incubated at 99 °C in a waterbath for another 10 min to inactivate the proteinase. Samples were vortexed, spun at 3,000 g for 1 h in 50 mL conical tubes, and supernatant was trans-
ferred to a new tube and spun again at 15,000 g at 3,000 °C for 10 min. Thereafter, samples were incubated at 99 °C in a waterbath for another 10 min to inactivate the proteinase. Samples were vortexed, spun at 3,000 g for 1 h in 50 mL conical tubes, and supernatant was trans-
ferred to a new tube and spun again at 15,000 g at 3,000 °C for 10 min. Thereafter, samples were incubated at 99 °C in a waterbath for another 10 min to inactivate the proteinase. Samples were vortexed, spun at 3,000 g for 1 h in 50 mL conical tubes, and supernatant was trans-
ferred to a new tube and spun again at 15,000 g at 3,000 °C for 10 min. Thereafter, samples were incubated at 99 °C in a waterbath for another 10 min to inactivate the proteinase. Samples were vortexed, spun at 3,000 g for 1 h in 50 mL conical tubes, and supernatant was trans-
ferred to a new tube and spun again at 15,000 g at 3,000 °C for 10 min. Thereafter, samples were incubated at 99 °C in a waterbath for another 10 min to inactivate the proteinase. Samples were vortexed, spun at 3,000 g for 1 h in 50 mL conical tubes, and supernatant was trans-
ferred to a new tube and spun again at 15,000 g at 3,000 °C for 10 min. Thereafter, samples were incubated at 99 °C in a waterbath for another 10 min to inactiva-
same time into identical tubes from healthy volunteer donors (following informed consent, under the approval from the University of California, San Diego Human Subjects Institutional Review Board), and stored overnight on ice, to ensure similar treatment conditions before analysis. All health and safety issues related to handling of human and nonhuman primate samples are covered by an Institutional biosafety approval from the University of California, San Diego Environmental Health and Safety Committee. All individuals who handle the samples receive the required training regarding precautions for blood-borne pathogens.

Mice and Chow. Cmah−−/− mice were described previously (26) and syngeneic WT C57BL/6 mice were purchased from Harlan Laboratories. Mice were fed a vegan standard chow (110951/110751; Dyets, Inc.) or WTC57BL/6 mice were purchased from Harlan Laboratories. Mice were fed in accordance with The Association for Assessment and Accreditation of Laboratory Animal Care under protocol S01227 approved by The Institutional Animal Care and Use Committee of the University of California, San Diego.

Isolation of GAGs. An adult mouse liver and kidney were each homogenized in 2 mL and 1 mL of 0.1x PBS, pH 7.4 (BioPioneer) containing 1:100 protease inhibitor mixture III (Calbiochem) using a polytron, respectively. The resulting kidney and liver homogenates were diluted with buffer A to a final volume of 10 mL, whereas only 500 μL of the liver homogenate were used and diluted to 20 mL with buffer A. For purification of GAGs from animal muscle, 2 g of ground tissue were homogenized with 0.01% sodium azide and 100 μL of the liver homogenate were incubated in a head-over-head shaker at 37 °C overnight. The sample was lyophilized and CS disaccharides tagged with [12C6]aniline as described previously (23, 28). In brief, 130 mg NaNbH₄ were dissolved in 1.4 mL DMSO before adding 0.6 mL glacial acetic acid. Enzymatically depolymerized GAG samples were dried down after digestion and reconstituted in 17 μL [12C6]aniline to which 17 μL of the above-mentioned NaNbH₄ solution were added. The reactions were incubated for 16 h at 37 °C (dry heat). Thereafter, samples were dried down using a speedvac and stored in the dark at 4 °C until analyzed by LC/MS.

Release of GAGs from Fossils and Bones. A rotary power tool (Dremel) with a brush attachment was used to carefully remove debris from the outside surface of the fossils and bones; and Jeffrey D. Esko for his critical review of the manuscript. This work was supported by NIH Grants 1R13GM073273 (to A.V.), GM33063, and HL107150 (J. Esko, Principal Investigator) and the Mathers Foundation of New York. Mass spectrometry analyses were performed at the Glycotechnology Core Resource at the Glycobiology Research and Training Center, University of California, San Diego, managed by Biswa Choudhury.

4. Bergfeld AK, Pearce OM, Daza SL, Pham T, Varki A (2012) Metabolism of vertebrate amino sugars with N-glycolysed CS (50 μg) described above was incubated with 50 mM Tris HCl, pH 7.9, 50 mM NaCl, and 20 μL of chondroitinase ABC (Sigma) in a final volume of 50 μL at 37 °C overnight. The sample was lyophilized and CS disaccharides tagged with [12C6]aniline as described above but using [13C6]aniline instead of [12C6]aniline.

ACKNOWLEDGMENTS. We thank Eytty Indriadi and the late Teuku Jacob, Gadjah Mada University, Yogyakarta, Indonesia for the Indonesian Homo samples; the Yerkes Primate Research Center for the chimpanzee serum samples; healthy volunteers for blood samples; David Vocadlo for the UDP-HexNGc standard used in previous related studies; Pamela Stanley for CHO ic.eC29.LEC32 cells; Margaret Schoeninger from the Center for Academic Affairs, and Training in the Geological Sciences for the carbon, oxygen, and nitrogen isotope analyses; and Jeffrey D. Esko for his critical review of the manuscript. This work was supported by NIH Grants R01GM32373 (to A.V.), GM33063, and HL107150 (J. Esko, Principal Investigator) and the Mathers Foundation of New York. Mass spectrometry analyses were performed at the Glycotechnology Core Resource at the Glycobiology Research and Training Center, University of California, San Diego, managed by Biswa Choudhury.


Fig. S1. Synthesis of partially N-glycolylated CS. (A) Schematic representation of the procedure to generate partially N-glycolylated CS from commercial-grade shark cartilage CS. (B) The extracted ion current chromatograph for [13C6]aniline-labeled nonmodified singly sulfated disaccharide residues present after chemical modification. The relative abundances of both D0a4 and D0a6 are consistent with unmodified starting material. (C) The extracted ion current chromatograph for the corresponding [13C6]aniline-labeled glycolylated disaccharides D0q4 and D0q6. (D) The product ion mass spectrum for D0q4. (E) The product ion mass spectrum for D0q6.
Fig. S2. Disaccharide compositional analyses of CS. Relative abundance of CS disaccharides in contemporary (human and chimpanzee bone) compared with fossil specimens expressed as molar percent (mol %).
Fig. S3. Disaccharide compositional analyses of HS. Relative abundance of HS disaccharides in contemporary (human and chimpanzee bone) compared with fossil specimens expressed as molar percent (mol %).
<table>
<thead>
<tr>
<th>CS purified from</th>
<th>Presence of Gc-CS</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT mouse liver</td>
<td>Yes (Fig. 3)</td>
</tr>
<tr>
<td>Cmah&lt;sup&gt;−/−&lt;/sup&gt; mouse liver</td>
<td>No (Fig. 3)</td>
</tr>
<tr>
<td>WT mouse kidney</td>
<td>Yes</td>
</tr>
<tr>
<td>Cmah&lt;sup&gt;−/−&lt;/sup&gt; mouse kidney</td>
<td>No</td>
</tr>
<tr>
<td>WT mouse serum</td>
<td>Yes (Fig. 4)</td>
</tr>
<tr>
<td>Cmah&lt;sup&gt;−/−&lt;/sup&gt; mouse serum</td>
<td>No (Fig. 4)</td>
</tr>
<tr>
<td>Bovine muscle (beef)</td>
<td>Yes</td>
</tr>
<tr>
<td>Porcine muscle (pork)</td>
<td>Yes</td>
</tr>
<tr>
<td>Ovine muscle (lamb)</td>
<td>Yes (Fig. 3)</td>
</tr>
<tr>
<td>FCS</td>
<td>Yes</td>
</tr>
<tr>
<td>Bovine trachea (commercial CS)</td>
<td>Yes</td>
</tr>
<tr>
<td>Porcine intestinal mucosa (commercial CS)</td>
<td>Yes</td>
</tr>
<tr>
<td>Chimpanzee serum</td>
<td>Yes (Fig. 4)</td>
</tr>
<tr>
<td>Human AB serum</td>
<td>Trace (Fig. 4)</td>
</tr>
<tr>
<td>Human serum (S-8)</td>
<td>Trace</td>
</tr>
<tr>
<td>Human serum (S-19)</td>
<td>Trace</td>
</tr>
<tr>
<td>Human serum (S-30)</td>
<td>Trace</td>
</tr>
<tr>
<td>Human serum (S-34)</td>
<td>Trace</td>
</tr>
<tr>
<td>Human serum (S-39)</td>
<td>No</td>
</tr>
<tr>
<td>Human serum (S-46)</td>
<td>No</td>
</tr>
<tr>
<td>Human serum (S-54)</td>
<td>Trace</td>
</tr>
<tr>
<td>Human serum (S-55)</td>
<td>Trace</td>
</tr>
<tr>
<td>Human serum (S-56)</td>
<td>Trace</td>
</tr>
<tr>
<td>Cmah&lt;sup&gt;−/−&lt;/sup&gt; mouse serum from Neu5Gc-fed mice</td>
<td>Trace (Fig. 4)</td>
</tr>
<tr>
<td>Shark cartilage (commercial CS)</td>
<td>No (Fig. S1)</td>
</tr>
</tbody>
</table>

Summary of samples analyzed for the presence of N-glycolylated CS.
Table S2. Detection of N-glycolyl-CS in fossil specimens

<table>
<thead>
<tr>
<th>Name</th>
<th>Origin</th>
<th>Location</th>
<th>Estimated age</th>
<th>Description/site where collected</th>
<th>Total amount of CS, μg/g</th>
<th>Total amount of HS, μg/g</th>
<th>N-glycolyl-CS detectable?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>Human (humerus)</td>
<td>USA</td>
<td>Contemporary</td>
<td>Provided by USCD Anatomical Preparation</td>
<td>139.2</td>
<td>337.8</td>
<td>Yes (*)</td>
</tr>
<tr>
<td>Chimpanzee</td>
<td>Chimpanzee (vertebrae)</td>
<td>USA</td>
<td>Contemporary</td>
<td>Provided by CARTA</td>
<td>68.1</td>
<td>657.4</td>
<td>Yes (***)</td>
</tr>
<tr>
<td>F16</td>
<td>Mammoth (astragalus)</td>
<td>Dickinson Ct, Kansas, USA</td>
<td>~12,000 y</td>
<td>Forals Fossils (Alan Foral, 5417 Calvert, Lincoln, NE 68506)</td>
<td>2.6</td>
<td>11.1</td>
<td>Yes (*)</td>
</tr>
<tr>
<td>F19</td>
<td>Elk (antler)</td>
<td>Dickinson Ct, Kansas, USA</td>
<td>~12,000 y</td>
<td>Forals Fossils (Alan Foral, 5417 Calvert, Lincoln, NE 68506)</td>
<td>28.8</td>
<td>106.4</td>
<td>Yes (****)</td>
</tr>
<tr>
<td>F20</td>
<td>Cave bear (jaw bone)</td>
<td>Romania</td>
<td>40,000–80,000 y</td>
<td>Canada Fossils (John Issa, 3333 Eighth Street SE, Calgary, Alberta T2G3A4)</td>
<td>4.6</td>
<td>178.7</td>
<td>Yes (++++)</td>
</tr>
<tr>
<td>F71</td>
<td><em>H. Erectus</em> (Sangiran 27, skull)</td>
<td>Sangiran, Java, Indonesia</td>
<td>1.65 My</td>
<td>Teuku Jacob, Etty Indriati</td>
<td>0.2</td>
<td>86.1</td>
<td>n.d.</td>
</tr>
<tr>
<td>F33</td>
<td>Bovidae</td>
<td>West Turkana Kenya</td>
<td>1.47 My</td>
<td>Nariokotome 3 (NK3) next to excavation of KNM-WT 15,000, <em>H. erectus</em> skeleton</td>
<td>0.054</td>
<td>n.a.</td>
<td>n.d.</td>
</tr>
<tr>
<td>F34</td>
<td>Bovidae (mandible fragment, WT 4608)</td>
<td>West Turkana, Kenya</td>
<td>Between 0.7 and 1.4 My</td>
<td>Nachukui 1 (NC1) above Chari Tuff</td>
<td>0.089</td>
<td>n.a.</td>
<td>Yes (++++)</td>
</tr>
<tr>
<td>F42</td>
<td>Bovid or hippo (occipital condyle fragment)</td>
<td>West Turkana, Kenya</td>
<td>3.4 My</td>
<td>Lomekwi 4 (L04), level of Tulu Bor tuff</td>
<td>0.013</td>
<td>n.a.</td>
<td>Yes (+)</td>
</tr>
<tr>
<td>F44</td>
<td>Mammal (vertebra fragment)</td>
<td>West Turkana, Kenya</td>
<td>3.4 My</td>
<td>Lomekwi 4 (L04), level of the Tulu Bor tuff</td>
<td>0.026</td>
<td>n.a.</td>
<td>Yes (+++)</td>
</tr>
<tr>
<td>F56</td>
<td>Mammal (long bone fragment)</td>
<td>Allia bay, Kenya</td>
<td>3.9 My</td>
<td>Collected in excavation 261-1 with <em>Australopithecus anamensis</em></td>
<td>0.051</td>
<td>n.a.</td>
<td>Yes (+)</td>
</tr>
</tbody>
</table>

Total amounts of CS and HS were calculated by MS using a [13C6]aniline-tagged internal standard. The presence of N-glycolyl-CS in all samples is provided with the number of crosses and asterisks indicating relative abundance of D0q4 among the two sample sets. n.a., not analyzed; n.d., not detectable.